Oslo Business School
Working Paper 4/2017

Price Dispersion and the Role of Stores

Espen Moen, Fredrik Wulfsberg and Øyvind Aas

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Price Dispersion and the Role of Stores*

Espen Moen
BI Norwegian Business School

Fredrik Wulfsberg
Oslo Business School

Øyvind Aas
Université libre de Bruxelles

August 9, 2017

Abstract

This paper studies price dispersion in the Norwegian retail market for 766 products across 4 297 stores over 60 months. Price dispersion for homogeneous products is significant and persistent, with a coefficient of variation of 37% for the median product. Price dispersion differs between product categories and over time. Store heterogeneity accounts for 30% of the observed variation in prices for the median product-month and for around 50% for the sample as a whole. Price dispersion is still prevalent after correcting for store heterogeneity.

Keywords: Price dispersion, retail prices, store heterogeneity.
JEL: D2, D4, E3.

*We are grateful for comments from Birthe Larsen, Tore Nilssen, Magnus Söderberg by seminar participants at the 2014 Annual Meeting of the Norwegian Association of Economists, the 8th Nordic Summer Symposium in Macroeconomics, the 9th Nordic Workshop in Industrial Organization at the University of Oslo, the 29th EEA-ESEM meeting at Toulouse School of Economics, Norwegian School of Economics, BI Norwegian Business School, Statistics Norway and Oslo Business School. Øyvind gratefully acknowledges the financial support of the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013 / ERC grant agreement no. 339950). Corresponding author: Fredrik Wulfsberg, Oslo Business School, Oslo and Akershus University College of Applied Sciences, frewul@hioa.no

It is common knowledge that the price for a particular product or service may vary substantially between stores or outlets. One explanation for price dispersion, is that stores are different. Stores can be heterogeneous in a multitude of ways such as location, opening hours, parking facilities (see e.g. Dixit and Stiglitz (1977) and Weitzman (1994)), loyalty programs (Basso et al., 2009) and warranties (Grossman, 1981). In addition, idiosyncratic shocks or unexpected fluctuations in demand (Mackowiak and Wiederholt, 2009) may also yield price dispersion. Furthermore, store characteristics are often an intrinsic component of a purchase (product differentiation). For example, buying a lukewarm Coca-Cola in a supermarket in the middle of the day is different from buying a cold one from a convenience store or a petrol station in the middle of the night. Or, eating the same meal at two different restaurants may be perceived as very different depending on characteristics for each restaurant. Store characteristics may thus reflect different mark-ups and costs, and result in price dispersion.

In this paper we identify the contribution of store characteristics to price dispersion exploring monthly price observations from a wider set of products categories than previous studies. But first we establish four stylized facts of price dispersion: (1) there is significant and persistent price dispersion in retail prices in Norway, the median standard deviation is 33% of the mean price. (2) The dispersion of prices varies between products and over time as indicated by the inter quartile rage of the standard deviation between 19% and 50%. (3) 84% of the overall variation in the standard deviation is between products while 16% is due to time variation. (4) The dispersion in prices increased from around 25% in the start of the sample to almost 40% in end of the sample.

Second, we identify a fixed store component of prices by observing prices of multiple products in the same store over time. Using intuitive non-parametric methods we find that the store component accounts for about 30% of the price dispersion *for the median product.* To identify the store component for the sample as a whole we also employ a novel parametric method, which shows that store effect accounts for 50–60% of the dispersion in prices. As further evidence of the importance of store heterogeneity, we find that the ranking of stores within the price distributions is highly persistent over time.

Kaplan and Menzio (2015) use scanner data from 1.4 million (mostly food) products across different geographical areas in the United States. They find that the quarterly average standard deviation in prices is between 19 and 36% depending on the aggregation level,[1] and that store heterogeneity account for 10% of the dispersion in prices. Exploring a subset of the data and a different method, Kaplan et al. (2016) find that 15.5% of the variance in prices is due to store heterogeneity.[2] Lach (2002) studies price dispersion for only four products[3] in Israel. He finds that store characteristics account for between 47–90% of the variation in prices. Wildenbeest (2011) investigate price dispersion of a basket of grocery items from four retailers in the United Kingdom. He finds that store heterogeneity explains around 61% of variation in prices and attributes the rest to search frictions.

We contribute to this literature by covering a larger variety of products from multiple stores. We include not only food products, but also products from all 10 COICOP categories such as consumer electronics, cars, petrol, apparel, restaurants, transport, and other services. This allows us to provide more detailed insight into price dispersion than previous studies. Also, we argue that the store effect may represent information about the price structure in the market.

1 The role of the stores

Our data covers hundreds of products from most product categories and from thousands of stores. However, households may choose from a tremendously large

[1] Products are defined by universal product code (UPC).

[2] They also decompose the variation in prices into transitory and persistent parts. The persistent component of the store-product variation in prices (which they label "relative price dispersion") constitutes 30.3% of the variation in prices.

[3] These are refrigerator, chicken, flour, and coffee.

set of products of different brands and qualities at different prices from different stores. Considering what products to buy where is a huge task for consumers, and it is impossible in practice to gather and process all available information. In such an environment, searching for products and stores may easily become random and price competition weak. Changing a price of a product will have small demand effects, as the probability that the consumers notice it is rather small. In this case price dispersion indicates inefficiencies and costs.

However, stores may be helpful for consumers by reducing this information problem. Suppose the prices of different products within a store is perfectly correlated. Suppose also that the average price of a store relative to other stores is constant. Then the consumer only needs to know the average price in each store to be perfectly informed (actually, only the price of one good, as prices are perfectly correlated).[4] If, in addition, the average price distribution is reasonably stable, the information problem facing consumers is reduced to something that is rather simple. In this case, the stores will compete through having low average costs, and there is no reason why price competition should not be hard.

Generally, it is much easier to know the average price in a store than knowing the price of each product. Surveys for instance, give information about average prices. Hence, through their price policy, stores may reduce the information problem facing the consumers, from chaos (no store effect) to small (strong store effect), with increased store effect going hand in hand with increasing price competition. The more is explained by the store effect, the easier is it for the customers to be informed about relative price levels. Hence measuring the store effect gives information about how easy it is to get information for customers, which again is a stepping stone for understanding the working of the price system as a whole. The store effect may thus be an indication of the information structure in the market, and a high store effect makes it more likely that the consumers are well informed

[4]Furthermore, if stores in the same chain have similar prices, it is sufficient for a consumer to know average prices in each chain in order to be perfectly informed.

and that competition works well.

Besides providing information of relative prices, store effects may also reflect other aspects of price setting as mentioned in the introduction. For example location and opening hours may yield differences in mark-ups or marginal costs between stores. A third explanation for store effects is product aggregation. As some products in our data may represent different brands and qualities across stores, this may yield a store effect. In our empirical analyses below we do not distinguish between the different sources of the store effect.

Obviously, one may ask why don't the store effect explain all dispersion in prices? First, it may be that markups are different for different products in the same store. Stores may for example lower prices on a few products to lure customers to shop at their outlet. It may also reflect cost differences of the individual products within the stores. Other stores may specialize in expertise to help consumers choose between products of different qualities and properties while selling more standardized products at market prices. For example a store selling cross country skis may specialize in skills finding a pair of skis with the appropriate span and stiffness and thus charge a higher price while charging the same price a general store for ski wax.

Also, incomplete information may yield a search equilibrium where prices differ randomly between stores cf. Burdett and Judd (1983) and Moen (1999).[5] Lastly, Kaplan and Menzio (2015) shows that price discrimination also may yield price dispersion beyond store effects.

2 Data

We use monthly micro data collected for the consumer price index (CPI), see Statistics Norway website[6] and Statistics Norway (2001). The data covers monthly

[5]Moen (1997) shows that price dispersion also may emerge with directed search.

[6]`http://www.ssb.no/en/priser-og-prisindekser/statistikker/kpi/`

price observations of 760 different products and services from 4,297 stores from January 2000 to December 2004. In total, our sample consists of 2,774,494 price observations.

The products represent all COICOP[7] categories such as food, apparel, furnishing, transport, services, recreation, electronics, and fuels to name a few. Compared to Kaplan and Menzio (2015) and Kaplan et al. (2016) our data includes a larger variety of products covering more than 70% of household expenditures. The panel is unbalanced since some products and stores are replaced each year by Statistics Norway to ensure the representativeness of the consumption basket (Statistics Norway, 2001). The products are defined with varying degree of precision depending on its type. For example, *Coca-Cola, bottle, 0.33 liters* is more precise than *Dress, ladies, simple*. Thus, for some products price dispersion may reflect differences in quality between stores. Appendix A reports some descriptives on the size of the data set.

3 Stylized facts of price dispersion

In this section, we present different measures of price dispersion in our sample and how dispersion varies between products and over time. We denote P_{ist} as the price observation for product i in store s at month t. First, we construct a price distribution for each product-month sample, in total 40,567 distributions. We drop product-months with fewer than 20 observations (stores) in order to reduce sampling errors. For example, Figure 1 shows the price distribution of observed prices P_{ist} for a bottle of Coca-Cola (in NOK) from 274 stores in January 2000. The lowest price is 7 NOK and the highest price more than three times higher at 24 NOK. The third quartile price is 89% higher than the first quartile price, and the standard deviation is 3.70 NOK which is 28% relative to the mean price.

[7] COICOP is an acronym for Classification of Individual Consumption According to Purpose, which is a nomenclature developed by the United Nations Statistics Division to classify and analyze individual consumption expenditures incurred by households according to their purpose.

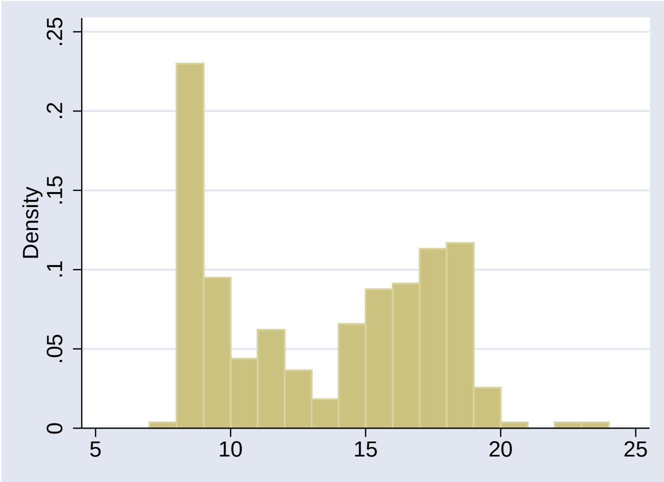

Figure 1: The distribution of prices in NOK *of Coca-Cola in January 2000.*

Interestingly, the price distribution is clearly bimodal. One possible interpretation of this feature is that stores are either "cheap" (like supermarkets) or "expensive" (like convenience stores).

Each product-month distribution of P_{ist} will obviously have different means μ_{it} and variances Σ_{it} which may depend on the the scale (i.e. the mean μ_{it}). In order to compare the dispersion of prices across products and months we normalize all prices P_{ist} with respect to the mean price for each product in each month $\bar{P}_{it}$:

$$\tilde{P}_{ist} = P_{ist}/\bar{P}_{it}$$

$\tilde{P}_{ist}$ has thus a mean of 1 and variance σ^2_{it}, which we can compare across products and over time.[8]

As the product-month price distributions may be skewed or multimodal like the distribution for Coca-Cola in Figure 1, we measure price dispersion as the interquartile range (IQR) and the 95/5 percentile range (P95–P5 range) in addition to the variance and standard deviation. Since the distributions are normalized, the standard deviation, the IQR and the P95–P5 range are in percent of the mean.

The first column of Table 1 reports the median product-month estimates of

[8] We exclude outliers with a relative price greater than 5 or less than 0.05 and then renormalize.

Table 1: Descriptive statistics for relative price dispersion across products and over time.

Dispersion measure	Median	$(Q1-Q3)$ range
Standard deviation	0.327	(0.193 – 0.504)
IQR	0.319	(0.166 – 0.562)
$(P95-P5)$ range	0.945	(0.548 – 1.480)

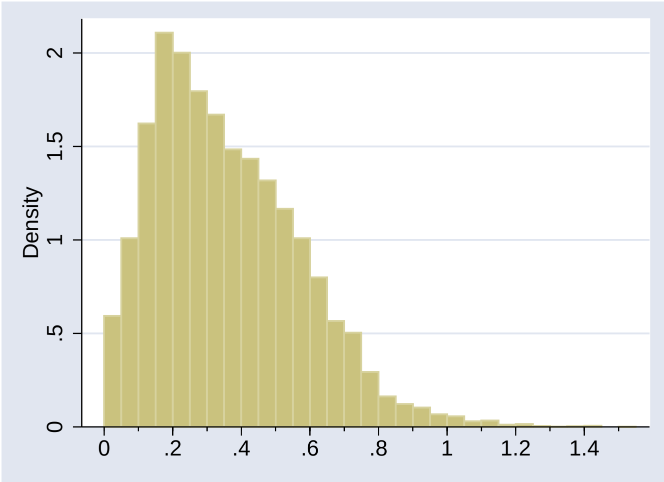

Figure 2: The histogram of the standard deviation for all product-month distributions truncated at 2.

these measures, and in column 2 we illustrate the variation in each of these measures (across products and over time) by the range between the first and third quartile, i.e. the $(Q1-Q3)$ range. Table 1 shows that price dispersion is significant with a estimated median standard deviation of 32.7% (corresponding to a variance of 0.107). The median IQR is 31.9% and the median $(P95-P5)$ range is 94.5%. However, there is a lot of variation between the product-month distributions as indicated by a $(Q1-Q3)$ range of the standard deviation between 19% and 50%. Similarly there is a lot of variation in the other measures of price dispersion. The $(Q1-Q3)$ range for IQR is between 16.6% and 55.8%, and between 54.8% and 149.0% for the $(P95-P5)$ range. This variation may reflect that some products are precisely defined while others are aggregates of products that are close substitutes. Figure 2 shows a histogram of the 40,567 standard deviations illustrating

Table 2: Median dispersion in relative prices across COICOP *categories. Ranked by median standard deviation.*

		Standard deviation	
COICOP division	N	Median	$(Q1-Q3)$ range
3 Clothing and footwear	102	0.551	(0.425 – 0.666)
5 Furnishings, household equip	120	0.424	(0.254 – 0.584)
11 Restaurants and hotels	38	0.408	(0.307 – 0.488)
8 Communication	7	0.401	(0.343 – 0.523)
7 Transport	37	0.351	(0.161 – 0.515)
9 Recreation and culture	81	0.325	(0.199 – 0.488)
12 Miscellaneous goods, services	60	0.325	(0.219 – 0.536)
1 Food and non-alcoholic bevs	255	0.253	(0.170 – 0.367)
4 Housing, water, electricity, gas and other fuels	15	0.250	(0.134 – 0.385)
2 Alcoholic beverages, tobacco and narcotics	12	0.099	(0.068 – 0.179)
6 Health	39	0.089	(0.035 – 0.223)
Semi-durables	184	0.508	(0.384 – 0.644)
Services	72	0.388	(0.248 – 0.497)
Durables	86	0.372	(0.213 – 0.553)
Non-durables	424	0.243	(0.154 – 0.376)

Note: N is number of products in each category.

the variation in price dispersion. The distribution is skewed right which is why we report on the median.[9]

Grouping the products by COICOP categories shows systematic differences in dispersion between types of products. Table 2 shows the median standard deviation and IQR for each COICOP division in the top panel and the degree of durability in the bottom panel. The groups are ranked according to their standard deviation. *Clothing and footwear* products have the highest price dispersion with a median standard deviation of 55.1%. The least dispersed categories are *Health* products and *Alcoholic beverages and tobacco* with a median standard deviation of less than 10%. In particular, the dispersion we find in normalized prices for food products

[9] The mean variance and standard deviation are 0.180 and 36.3%.

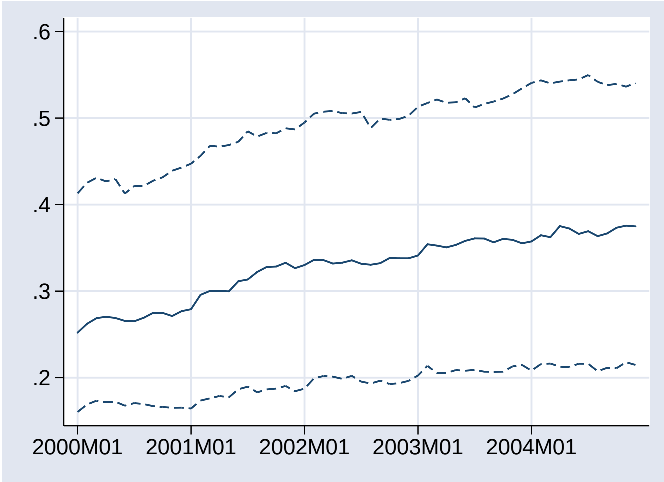

Figure 3: The first, second (median) and third quartiles of the standard deviations over time.

in Norway is 25.3% (measured by the median standard deviation), which is less than what Kaplan and Menzio (2015) find for food products in the United States (36%).

The COICOP system also classifies products as durables, semi-durables, non-durables and services. For example, clothing and footwear is classified as semi-durables. The bottom part of Table 2 shows the median product-month standard deviation within these categories. The standard deviation for semi-durable products is about twice as high as for non-durable products (50.8% vs. 24.3%). The dispersion of services and durables are in between at around 38%. The right column of Table 2 shows that there is a lot of heterogeneity within each consumption category indicated by the $(Q1 - Q3)$ range of the standard deviation.

Figure 3 shows that there is a clear upward trend in the three quartiles (Q1, median and Q3) of the standard deviation indicating that dispersion increased over time. The median standard deviation increased from around 25% in the beginning of the sample to almost 40% in the end. This finding is consistent with Wulfsberg (2016) who reports that the mean size of nonzero price changes increased over the same period.

How much of the variation in price dispersion reported in Table 1 is accounted for by this trend? Decomposing the variation in the standard deviation σ_{it} into cross sectional variation between products ($\bar{\sigma}_i$) and time variation within products ($\sigma_{it} - \bar{\sigma}_i$) yields that the cross sectional variation accounts for 84% while time variation only accounts for 16% of the overall variation in σ_{it}. 0What can explain this trend? In a Ss-type menu cost model the size of price adjustments increase with inflation. Hence, we would expect more price dispersion when inflation is high. However, during this period the 12 month annual inflation rate has varied between 4% and –2% with a slightly negative trend (if any).

We noted above that a possible explanation for the bimodality of the price distribution for Coca Cola in Figure 1 is that the product is sold by stores with different characteristics such as "cheap" stores (supermarkets) and "expensive" stores (convenience stores). Multimodal product-specific distributions of normalized prices seems to be prevalent. Instead of visual inspection we employ the Hartigan dip test (Hartigan and Hartigan, 1985) which rejects unimodality at 1% level of significance for as many as 576 (75%) of the products.[10] However, the pooled distribution of normalized prices is single peaked, kurtotic, and slightly left skewed as seen from Figure B1 in the appendix. The standard deviation of of the pooled distribution is 39.9% corresponding to a variance of 0.157 which is larger than the median product-month variance (0.107).

4 The store component

We assume that the relative price $\tilde{P}_{ist}$ can be decomposed into the mean (by definition equal to one), a store component v_s and a residual ε_{ist}:

$$\tilde{P}_{ist} = 1 + v_s + \varepsilon_{ist} \tag{1}$$

[10] Cavallo and Rigobon (2012) uses the dip test to inspect the distribution of price changes.

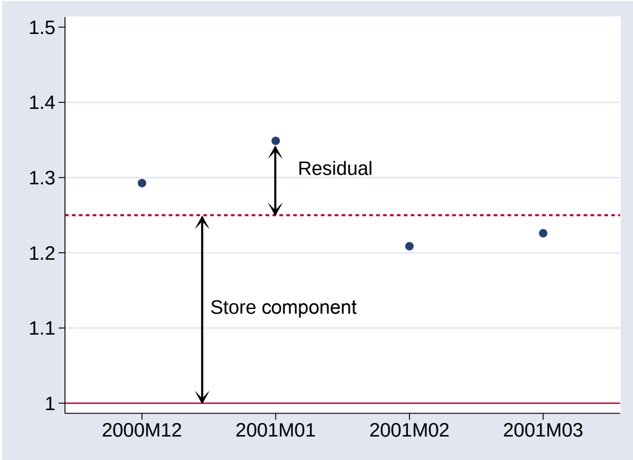

Figure 4: Illustration of the store component. The vertical axis is the relative price and the horizontal axis is the month. The dots are observations from the same store.

The store component is assumed to be equal for all products sold in the same store s in all periods. The idea of the store component is illustrated in Figure 4 where the dots represent four observations of the relative price for a Coca Cola sold in one particular store s from December 2000 to April 2001. The distance from the mean relative price (equal to 1) to the dashed line represents the store component, v_s. The conditional mean of the relative price for all products sold by this particular store, is thus $1 + v_s$. Since the store component is around .25 the store is on average 25% more expensive compared to other stores. However, the four price observations vary around the store mean $(1 + v_s)$, represented by the residual ε_{ist}.[11]

An intuitive estimate of the store components v_s is the mean normalized price for all products in all periods sold by the same store:

$$\hat{v}_s = \frac{1}{N_s} \sum_n (\tilde{P}_{ist} - 1) \tag{2}$$

where $n = 1, \ldots, N_s$ is the number of observations from store s over all products and months. The median number of products per store is 46 (the first quartile

[11] While the store effect in (1) is fixed over time, Kaplan et al. (2016) estimate a time varying store effect decomposing the error terms further into a transitory and a persistence component. Their results indicate that 95% of the sample store effect is persistent.

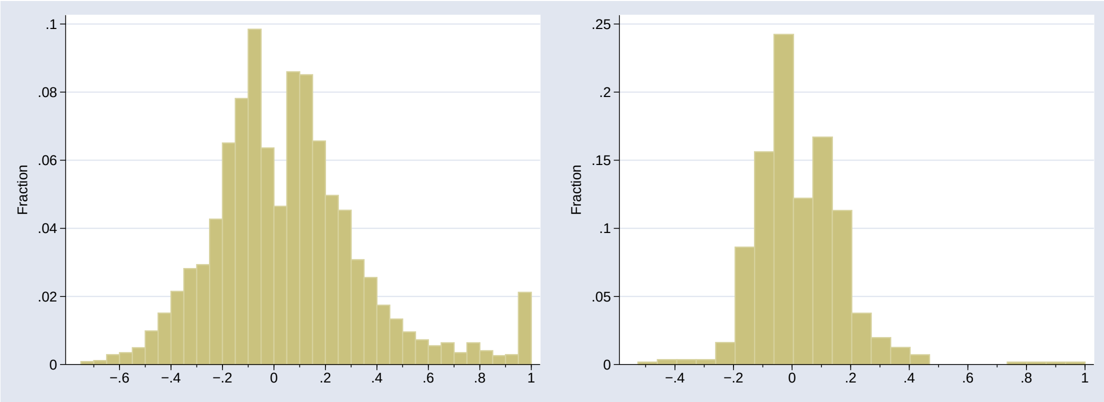

Figure 5: Histograms of the significant store effects (1% level significance) for all stores (left) and Coca-Cola stores (right). The histograms are truncated at 1.

number is 19 and the third quartile number is 187).

A t-test shows that one fifth of the store effects are insignificant at the 1% level.[12] The significant (non-zero) store effects are plotted in Figure 5. Their size vary between -70% and 300%. The histogram is clearly bimodal with one mode below zero around -10% and one mode above zero around 15%. The mean store effect is 28.5% for expensive stores and -18.0% for cheap stores. (The mean absolute size of the store effects is 23.6%.) The histogram to the right is for stores selling Coca-Cola which is also possibly bimodal with modes on each side of zero. In Figure C2 in the appendix the store effects are plotted by COICOP division. Bimodality is also reflected in these histograms with the exception of *9 Recreation and culture* and *12 Miscellaneous goods, services*. We note that the store effects seem particularly strong for *3 Clothing and footwear*.

Interestingly, there is a clear tendency that the more expensive the store is the more variation in (normalized) prices within the store. Figure 6 plots the store effects vs the variation in prices between products sold in the same store. One possible explanation for this is that expensive stores selling specialized products also need to sell standardized products (like our example above of a store selling cross-country skis and ski wax).

[12] 15% of the store effects are insignificant at the 5% level.

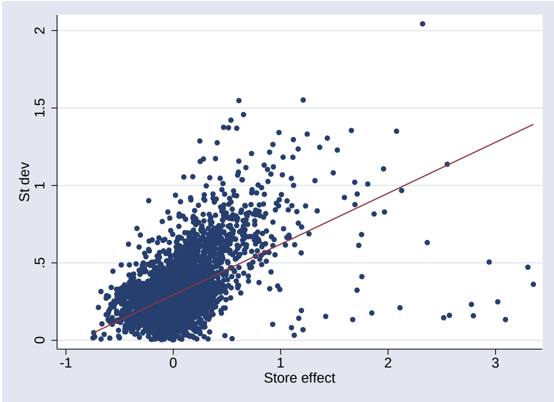

Figure 6: Expensive stores have higher price variation.

Table 3: Descriptive statistics for residual price dispersion across products and over time.

Dispersion measure	Median	$(Q1-Q3)$ range
Standard deviation	0.274	$(0.174-0.392)$
IQR	0.277	$(0.157-0.416)$
$(P95-P5)$ range	0.806	$(0.511-1.185)$

The residual component $\hat{\varepsilon}_{ist}$ is computed by subtracting the estimated store effects from each normalized price i.e. $\hat{\varepsilon}_{ist} = \tilde{P}_{ist} - 1 - \hat{v}_s$ following equation (1). The variance of ε_{ist} represents the price dispersion for products sold at equally expensive stores. Table 3 reports the same measures of residual price dispersion as for normalized prices in Table A1. Lach (2002) which focus on the dispersion of residual prices, finds less dispersion (but only for four products). All three dispersion measures of the residual prices are around 85% of the corresponding measure for normalized prices. There is thus substantial variation in prices even after controlling for store effects.

5 The importance of store heterogeneity

In this section, we explore how much of the variation in prices which we documented in the previous section, can be attributed to store heterogeneity. We interpret the statistical model (1) as an error component model where the store component v_s and the residual price ε_{ist} are stochastic each with a 0 mean, and a time and product specific variance σ^2_{vit} and $\sigma^2_{\varepsilon it}$. Note in particular that the variance of the store effect may vary between products and over time even if the store effect v_s does not vary between products and over time. The reason for this is that the product range varies between stores over time. To illustrate this point assume that there are in total three products A, B, and C. There are many stores, but each store sells only two products (A,B), (A,C) or (B,C). The store effect for a particular store is equal for both products sold in that store. However, the variance of the store effect for stores selling (A,B) products may be different from the variance of the store effect for stores selling the (A,C) products or the (B,C) products. Hence, the variance of the store effect for product A may differ from product B and product C.

Assuming that v_s and e_{ist} are independent, the variance of $\tilde{P}_{ist}$ is thus decomposed into

$$\sigma^2_{it} = \sigma^2_{vit} + \sigma^2_{\varepsilon it} \tag{3}$$

The ratio of the variance of the store component σ^2_{vit} to the total variance σ^2_{it} measures the importance of store heterogeneity for price dispersion. Our goal is thus to estimate σ^2_{vit} and $\sigma^2_{\varepsilon it}$ and their implied share of the total variance σ^2_{it}.

We first estimate the variance components for each product-month distribution based on the estimates $\hat{v}_s$ and $\hat{\varepsilon}_{ist}$ according to

$$\hat{\sigma}^2_{vit} = \sum_s (\hat{v}_s - \bar{v}_s)^2/(S_{it} - 1) \tag{4}$$

$$\text{and} \qquad \hat{\sigma}^2_{\varepsilon it} = \sum_s (\hat{\varepsilon}_{ist} - \bar{\varepsilon}_{ist})^2/(S_{it} - 1) \tag{5}$$

where $s = 1, \ldots, S_{it}$ is an index for stores selling product i in month t.

In column 1 of Table 4 we report the median variance of the store component

Table 4: Estimates of variance components. Share of total variance in parenthesis.

Variance:	Median		$(Q1-Q3)$ range	
store $\hat{\sigma}^2_{vit}$	0.032	(30)	(0.007–0.100)	(19–39)
residual $\hat{\sigma}^2_{eit}$	0.075	(70)	(0.030–0.154)	(81–61)
total $\hat{\sigma}^2_{vit}+\hat{\sigma}^2_{eit}$	0.107	(100)	(0.037–0.254)	(100)

and the median residual price variance, which turn out to be 0.032 and 0.075. This yields a total median variance of normalized prices of 0.107 (which is consistent with the estimates in Table 1). Thus the store effect accounts for 30% of the total variance of the median product-month. The ratio of the store variance varies obviously between product-month distributions. To illustrate this variation we report the same statistics for the inter quartile range (Q1–Q3 range) in the second column. We see that the store effect accounts for 19–39% of the total variance of product-month distributions measured by the Q1–Q3 range.

In Table 5 we report the variance decomposition by COICOP categories using the same approach as in Table 4. The store effect is particularly important for *11 Restaurants and hotels* accounting (49%) in addition to *3 Clothing and footwear* (45%). Typically for services we would expect variation in the store component to be an important part of the price dispersion. The store effect is least important for *8 Communication* with a ratio of 12% to the total variance. For food products (*1 Food and non-alcoholic beverages*) the store effects account for 23% of the total variance which is similar to Kaplan and Menzio (2015).

It is likely that stores selling the same product(s) are less heterogeneous than stores in general. For example, food stores are probably less heterogeneous than food stores versus hotels. Hence, for retail prices in general one may expect the variance in the store component to be more important for price dispersion than for the median product-month sample. To investigate this possibility we pool the sample and assume that $\sigma^2_{it}=\sigma^2$, $\sigma^2_{vit}=\sigma^2_v$ and $\sigma^2_{\varepsilon it}=\sigma^2_\varepsilon\ \forall\ i,t$. In this exercise

Table 5: Variance decomposition by COICOP division. Mean estimates.

COICOP division	σ^2	σ_v^2	σ_v^2/σ^2
11 Restaurants and hotels	0.151	0.074	49%
3 Clothing and footwear	0.333	0.150	45%
12 Miscellaneous goods, services	0.181	0.057	32%
2 Alcoholic beverages, tobacco and narcotics	0.055	0.017	32%
5 Furnishings, household equip	0.222	0.066	30%
4 Housing, water, electricity, gas and other fuels	0.098	0.026	26%
9 Recreation and culture	0.143	0.036	25%
6 Health	0.078	0.018	23%
1 Food and non-alcoholic bevs	0.109	0.025	23%
7 Transport	0.196	0.035	18%
8 Communication	0.189	0.022	12%
Services	0.159	0.060	38%
Semi-durables	0.293	0.106	36%
Durables	0.191	0.067	35%
Non-durables	0.112	0.027	24%

Table 6: Estimates of variance components. Pooled distribution

Variance:	Pooled		ME	
store $\hat{\sigma}_v^2$	0.104	(66)	0.098	(48)
residual $\hat{\sigma}_e^2$	0.053	(34)	0.107	(52)
total $\hat{\sigma}_v^2 + \hat{\sigma}_e^2$	0.157	(100)	0.205	(100)

we estimate σ_v^2 and σ_ε^2 by

$$\hat{\sigma}_v^2 = \sum_s (\hat{v}_s - \bar{v}_s)^2/(S-1) \tag{6}$$

$$\text{and} \qquad \hat{\sigma}_\varepsilon^2 = \sum_n (\hat{\varepsilon}_{ist} - \bar{\varepsilon}_{ist})^2/(N-1) \tag{7}$$

where $s = 1, \ldots, S$ is an index for all stores, and $n = 1, \ldots, N$ is an index for all observations in the sample. As reported in column 1 of Table 6 this yields an estimate of the variance of the store component of 0.104, which is significantly larger than the estimate in Table 4 as expected. This accounts for 66% of the pooled total variance leaving 34% for the residual variance, indicating a significant larger role for store effects when we look at the whole sample of products.

For robustness we estimate σ_v^2 and σ_e^2 directly using the mixed effects (ME) method (see Rabe-Hesketh and Skrondal, 2012 for details). This method estimates $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ simultaneously using maximum likelihood without first estimating the store component v_s, but assuming normality and homoskedasticity, i.e. $v_s \sim N(0, \sigma_v^2)$. This approach yields an estimate of the store component variance σ_v^2 equal to 0.098 which is a share of 48% of the total variance, see column 2 of Table 6. Note that the ME estimate of the total variance is larger than the pooled variance because of the normality assumption while the empirical (pooled) distribution is kurtotic (see Figure B1).

The different approaches thus yields an estimated share of the store effects from 30% for the median product-month sample to around 50% for the pooled sample. Our results thus attribute a somewhat stronger importance to store heterogeneity than Kaplan and Menzio (2015) who attribute 10%–36% of the observed price dispersion to store heterogeneity. The main reason for this is that we analyze for a wider product range and hence more heterogeneous stores.

5.1 Persistence

Store heterogeneity is an important component of the observed price dispersion, as documented above. In order to investigate the persistence of the store heterogeneity we inspect the ranking of stores within the price distributions over time following Lach, 2002. Is a store's ranking in the price distribution persistent as indicated by the estimated store effects?

For each product-month, we partition each price distribution by the three quartiles ($Q1$, median, and $Q3$) and assign each store into one of the four quartile bins $QB1_{it}$, $QB2_{it}$, $QB3_{it}$, and $QB4_{it}$. For how long does a store remain in the same quartile bin? Furthermore, how likely is it that a store which changes its nominal price, jump from one quartile bin to another or remain in the same part of the relative price distribution? If a store is systematically more expensive, consumers can learn this information and take advantage of price differences. If some consumers are informed about prices while others are not (Varian, 1980) shows that it is optimal for a store to randomize its price .

We are interested in the likelihood of a store moving from a quartile bin to another. For each observation, we construct four dummy variables y_{istj} (where $j = 1, 2, 3, 4$) indicating that the store s belongs to quartile bin QBj for product i in period t:

$$y_{istj} = \begin{cases} 1 & \text{if the price of product } i \text{ in store } s \text{ belongs to } QBj \text{ in period } t \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

The likelihood for store s of moving from quartile bin k to quartile bin j in the next month for product i is assumed to be

$$\gamma_{kji} = Pr\left[y_{istj} = 1 | \mathbf{y}_{is(t-1)}\right] = \Phi\left(\beta_{0is} + \sum_{k=1}^{4} \beta_{kis} y_{is(t-1)k}\right) \tag{9}$$

where $\mathbf{y}_{is(t-1)}$ is a vector of $y_{is(t-1)j}$. Φ is the cumulative distribution function for the standard normal distribution. We estimate the conditional likelihood of the Probit model with maximum likelihood.

For each product i the conditional likelihoods γ_{kji} gives us all the elements of the one-step transition probability matrix. Table 7 report the conditional one-step transition probability matrix 1 month ahead for the median product.[13] Each row represents the conditional probability of moving staying in the same bin or to either of the three other bins. For example, the median probability of moving from the first quartile bin $QB1$ to the second quartile bin $QB2$ in the next month

[13]See Table D1 in the appendix for the mean probabilities.

Table 7: 1 month transition probability matrix, normalized prices $\tilde{P}_{ist}$. Median estimates with standard errors in parenthesis. All observations.

	Destination bin			
Origin bin	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.885	0.079	0.015	0.006
	(0.096)	(0.061)	(0.030)	(0.053)
$QB2_t$	0.071	0.830	0.076	0.011
	(0.056)	(0.145)	(0.062)	(0.084)
$QB3_t$	0.010	0.080	0.831	0.069
	(0.023)	(0.059)	(0.152)	(0.109)
$QB4_t$	0.003	0.008	0.062	0.925
	(0.010)	(0.017)	(0.039)	(0.058)

Note: The rows does not sum to one since each element is the median value. However. the rows sum to one for each individual product.

Table 8: 12 months transition probability matrix, normalized prices $\tilde{P}_{ist}$. Median estimates with standard errors in parenthesis. All observations.

	Destination bin			
Origin bin	$QB1_{t+12}$	$QB2_{t+12}$	$QB3_{t+12}$	$QB4_{t+12}$
$QB1_t$	0.632	0.216	0.065	0.031
	(0.194)	(0.125)	(0.092)	(0.124)
$QB2_t$	0.174	0.506	0.206	0.056
	(0.098)	(0.186)	(0.134)	(0.109)
$QB3_t$	0.046	0.200	0.521	0.202
	(0.056)	(0.099)	(0.178)	(0.158)
$QB4_t$	0.019	0.047	0.183	0.734
	(0.036)	(0.068)	(0.082)	(0.129)

Note: See Table 7

is $\gamma_{12} = 7.9\%$.

We see that a store is most likely to stay in the same quartile bin in the next period since the probabilities along the diagonal are the largest varying from 83–93%. Given that a store moves from one bin quartile to another, it is most likely to move to an adjacent quartile bin. The closest elements to the diagonal vary from 6.2–8.0%. The probability of jumping two quartile bins e.g. from $Q3$ to $Q1$ ranges between 0.8–1.5%. A store is least likely to move for one tail to the other, with a probability ranging from 0.3-0.6%. The matrix is quite symmetric, but the upper elements are somewhat larger than the corresponding lower elements. This indicates that the likelihood of moving down from for example the third quartile bin to the first quartile bin γ_{31} is smaller than moving up from the first quartile bin to the third quartile bin γ_{13}. The transition probability matrix varies across products, as indicated by the standard deviations.

We find the same pattern when we estimate the 12 month transition probability matrix, see the median probabilities in Table 8.[14] Even 12 months ahead a store is most likely to remain in the same quartile bin than move to any other bin. The median probability of being in the same quartile bin in 12 months varies between 51% and 73% compared to 83–93% for the 1 month ahead estimates in Table 7.

A change in a store's ranking within the relative price distribution can happen as a result of not only changing its own price, but also if other stores have changed their price. It is interesting to know the transition probabilities conditional on the store changing its own nominal price. Table 9 reports the conditional transition probability matrix 1 month ahead. Still, the largest probabilities are found on the diagonal ranging from 60% to 79%. If a store do change its ranking following a nominal price change, it is most likely to move to an adjacent quartile, with probabilities ranging between 14.0–17.1%. These probabilities are roughly double compared to the corresponding unconditional probabilities. The probability of jumping two quartiles, for example from $QB1$ to $QB3$, ranges now between 3.8–5.6%, increasing with a factor of 4 compared to the unconditional probabilities. Finally, the median probability of moving between the tail bins are 1.7 and 2.5%. The standard deviations are larger than in the unconditional estimation, so there is more variation in the transition probability matrices when we condition on a nominal price change.

Our results suggest there are persistent patterns in the ranking of stores within

[14] See Table D2 in the appendix for the mean probabilities. Table D3 in the appendix reports the 6 month transition probability matrices for the median, the mean, and the standard deviation.

Table 9: 1 month transition probability matrix conditional on nominal price changes, normalized prices $\tilde{P}_{ist}$. Median estimates with standard errors in paranthesis.

	Destination bin			
Origin bin	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.722	0.162	0.056	0.025
	(0.156)	(0.086)	(0.070)	(0.116)
$QB2_t$	0.140	0.595	0.171	0.045
	(0.081)	(0.166)	(0.104)	(0.114)
$QB3_t$	0.043	0.158	0.602	0.166
	(0.056)	(0.075)	(0.164)	(0.127)
$QB4_t$	0.017	0.038	0.140	0.788
	(0.035)	(0.044)	(0.063)	(0.096)

Note: See Table 7

a distribution consistent with the finding that fixed store effects is an important component of variation in prices. Knowing the ranking of stores from a previous period may imply significant search cost savings for consumers since the previous ranking is a fair bet for the current ranking.

Fixed store effects are likely to be related to the persistence of relative prices. It is thus possible for consumers to learn what stores are cheaper on average. But how is the relative price mobility of equally expensive stores? To answer this question we control for the store effects and estimate transition probability matrices for the residual prices $\hat{\varepsilon}_{ist}$? Tables 10–12 report the unconditional transition probability matrices for the residual prices for 1 and 12 months, and 1 month conditional on a nominal price change.[15]

Comparing the diagonal elements of the residual price transition probability matrix in Tables 10 and 11 with the corresponding probabilities for relative prices in Tables 7 and 8, we see that are even higher for $QB1$ and $QB2$, but a little lower for $QB3$ and $QB4$. Relative prices for equally expensive stores are thus still very persistent. Lach (2002) finds more flexibility for residual prices than our results.

[15]See Table E1 and E2 in the appendix for the mean probabilities. Table E3 reports the 6 month transition probability matrices for the median, the mean, and the standard deviation.

Table 10: 1 month transition probability matrix, residual prices $\hat{\varepsilon}_{ist}$. Median estimates and standard errors in parenthesis. All observations.

	Destination bin			
Origin bin	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.907	0.069	0.011	0.007
	(0.076)	(0.043)	(0.020)	(0.013)
$QB2_t$	0.069	0.846	0.074	0.011
	(0.041)	(0.093)	(0.043)	(0.018)
$QB3_t$	0.011	0.077	0.844	0.064
	(0.021)	(0.044)	(0.093)	(0.039)
$QB4_t$	0.006	0.010	0.064	0.917
	(0.013)	(0.018)	(0.035)	(0.036)

Note: The rows does not sum to one since each element is the median value. However. the rows sum to one for each individual product.

Table 11: 12 months transition probability matrix, residual prices $\hat{\varepsilon}_{ist}$. Median estimates and standard errors in parenthesis. All observations.

	Destination bin			
Origin bin	$QB1_{t+12}$	$QB2_{t+12}$	$QB3_{t+12}$	$QB4_{t+12}$
$QB1_t$	0.676	0.199	0.064	0.043
	(0.147)	(0.083)	(0.062)	(0.048)
$QB2_t$	0.185	0.529	0.218	0.062
	(0.079)	(0.132)	(0.086)	(0.057)
$QB3_t$	0.062	0.204	0.517	0.191
	(0.060)	(0.087)	(0.136)	(0.092)
$QB4_t$	0.039	0.058	0.183	0.697
	(0.052)	(0.005)	(0.075)	(0.133)

Note: See Table 10

However, he is analyzing only four products and with only one product for each store which may lead to a biased store effect is prices are not perfectly correlated withuin each store.

We also measure the duration of being in a particular bin for stores. Figure 7 presents box-plots (across products) showing the fraction of different spells within each quartile bin. Most spells are typically between 1 to 3 months within either of the quartile bins. But there is also a huge fraction of products where stores remain

Table 12: 1 month transition probability matrix conditional on nominal price changes, residual prices $\hat{\varepsilon}_{ist}$. Median estimates with standard errors in paranthesis.

	Destination bin			
Origin bin	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.769	0.136	0.044	0.028
	(0.129)	(0.072)	(0.051)	(0.040)
$QB2_t$	0.130	0.661	0.146	0.043
	(0.072)	(0.133)	(0.078)	(0.048)
$QB3_t$	0.037	0.135	0.669	0.137
	(0.051)	(0.076)	(0.130)	(0.078)
$QB4_t$	0.024	0.035	0.121	0.809
	(0.036)	(0.043)	(0.068)	(0.104)

Note: See Table 10

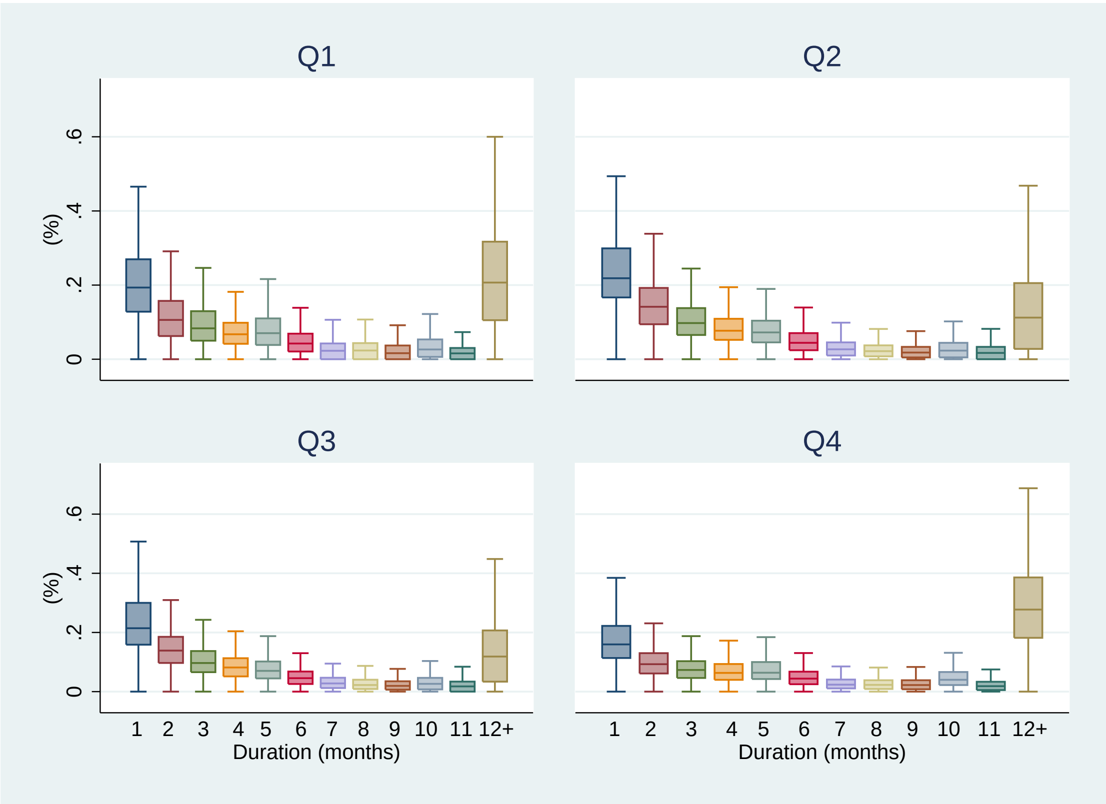

Figure 7: Box-plot of the monthly durations across the four quartiles. Note: See the appendix for the table with the data used to make the box plots. The table consists of mean, median and standard deviations.

in the same quartile bin for 12 months or more in particular the lower quartile bin $QB1$ and top bin $QB4$.

The relationship between the ranking spells and the transition probability matrix is the following. The conditional probability of changing to a different quartile is the sum of the off-diagonal elements in the transition probability matrix. Taking the average across the four quartiles, we get the probability of changing a quartile one month ahead. This probability is equal to the probability of observing a one--month spell (Lach, 2002). These probabilities are very similar for each individual product in our estimations.

Based on the ranking spells and the transition probability matrix, stores in our sample are persistently cheap or expensive. This result, combined with the result from the variance decomposition indicate that store heterogeneity is an important factor for price dispersion.

6 Conclusion

We document empirical facts of price dispersion for a wider range of retail products and services than in earlier studies. The standard deviation for the median product is 33%. Dispersion varies between products and months, indicated by the inter quartile range of the standard deviation from 19% to 50%. Prices for in particular clothing and footwear but also other semi-durable goods appear more dispersed than other products. Furthermore, price dispersion increased over time illustrated by an increase in the standard deviation for the median product from 25% to 40% over the sample period.

Our results suggest that store heterogeneity is an important component in price dispersion. By decomposing the variance in relative prices into a fixed store component and a idiosyncratic term, we find that 30% of the observed variance in relative prices for the median product-month can be account for by store heterogeneity. For the sample as a whole store heterogeneity accounts for 50% of the variance in relative prices, which is a larger share than reported in previous studies.

The distribution of the store components bimodal with a long right tail. The mean store effect for cheap stores is –18.0% while for expensive stores it is 28.5%.

The consistency of a stores ranking within a distribution indicate that most stores are likely to be in the same part of the distribution one month and even 12 months ahead.

From a consumer point of view, it is possible to learn what stores are cheap from searching for prices.

References

Basso, L. J., Clements, M. T., and Ross, T. W. (2009). Moral hazard and customer loyalty programs. *American Economic Journal: Microeconomics*, 1(1):101–123.

Burdett, K. and Judd, K. (1983). Equilibrium price dispersion. *Econometrica*, 51(4):955–969.

Cavallo, A. and Rigobon, R. (2012). The distribution of the size of price changes. *Working paper, MIT Sloan.*

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *The American Economic Review*, pages 297–308.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *Journal of law and economics*, 24:461–483.

Hartigan, J. and Hartigan, P. (1985). The dip test of unimodality. *The Annals of Statistics*, 13:70–84.

Kaplan, G. and Menzio, G. (2015). The morphology of price dispersion. *International Economic Review*, 56(4):1–41.

Kaplan, G., Menzio, G., Rudanko, L., and Trachter, N. (2016). Relative price dispersion: Evidence and theory. *NBER working paper series.*

Lach, S. (2002). Existence and persistence of price dispersion: An empirical analysis. *The Review of Economics and Statistics*, 84(3):433–444.

Mackowiak, B. and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, 99:769–803.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of Political Economy*, 105(2):385–411.

Moen, E. R. (1999). Education, ranking, and competition for jobs. *Journal of Labor Economics*, 17(4):694–723.

Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and longitudinal modmodel using stata. Volume I: Continuous Responses*. Stata Press.

Statistics Norway (2001). Consumer price index 1995–2000. *Official Statistics of Norway.*

Varian, H. (1980). A model of sales. *American Economic Review*, 70(4):651–659.

Weitzman, M. L. (1994). Monopolistic competition with endogenous specialization. *The Review of Economic Studies*, 61(1):45–56.

Wildenbeest, M. (2011). An empirical model of search with vertically differentiated products. *The RAND Journal of Economics*, 42(4):729–757.

Wulfsberg, F. (2016). Inflation and Price Adjustments: Micro Evidence from Norwegian Consumer Prices 1975–2004. *American Economic Journal: Macroeconomics*, 8(3):175–94.

Appendix

A Data dimensions

Table A1 presents descriptive statistics for the dimensions of our sample. The median number of observations per product over the sample period is 3,080, with an inter-quartile range (IQR) between 1,980 (Q1) and 4,751 (Q3) observations per product. About 2/3 of the products are observed over the whole period (60 months). While most products are observed over the entire period, no store is observed more than 47 months, with a median of 31 months. In each given month, there are 58 stores for the median product, and the median number of products in a store is 46 with and IQR between 19 an 187. There are 108 364 combinations of product and stores in the sample. See Wulfsberg (2016) for further descriptions of the data.

Table A1: Descriptive statistics.

	Median	$(Q1 - Q3)$ range
Observations per product	3,080	$(1,980 - 4,751)$
Number of months per product	60	$(53 - 60)$
Number of months per store	31	$(19 - 47)$
Number of stores per product-month	58	$(39 - 87)$
Number of products per store	46	$(19 - 187)$

Note: Q1 and Q3 are the first and third quartiles.

B The pooled distribution of normalized prices

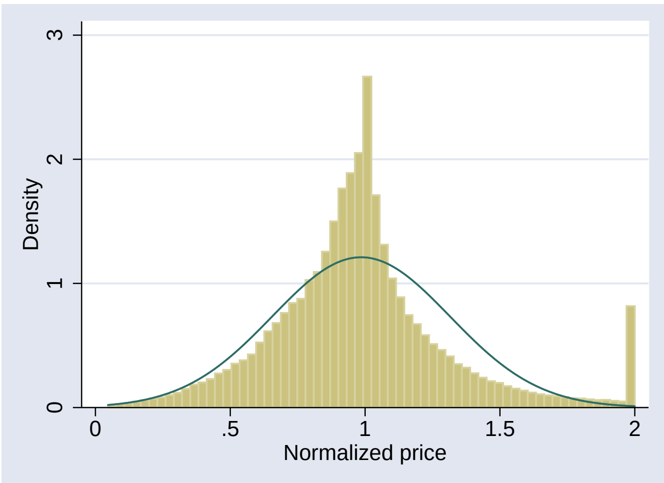

Figure B1: The pooled distribution of normalized prices truncated at 2.

C Store effects by COICOP division

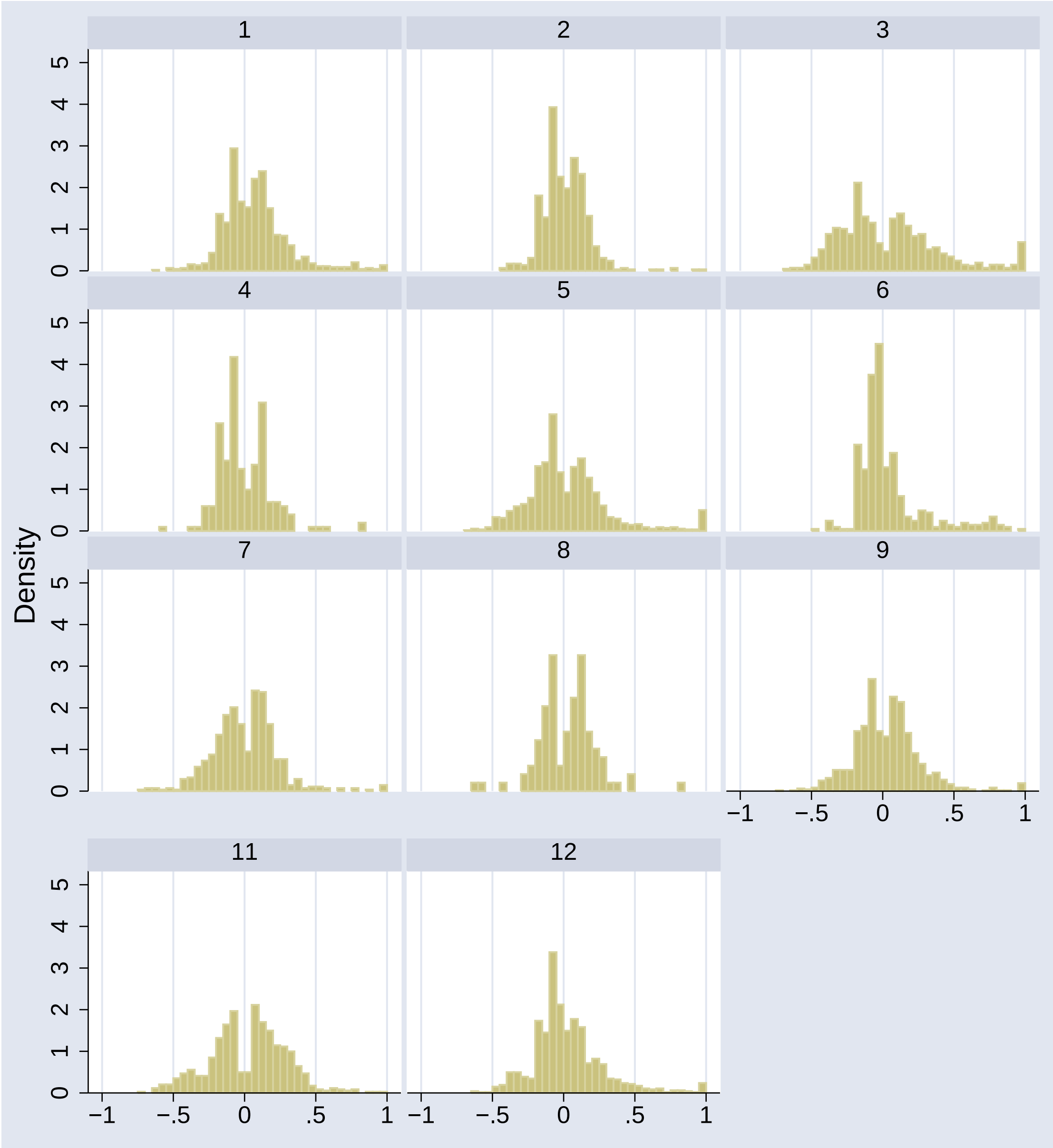

Figure C2: Histograms of the significant store effects (1% level significance) by COICOP division. The histograms are truncated at 1.

D Unconditional One Step Transition Probability Matrices

The elements in the matrix denote the probability of going from an initial quartile bin in period t (rows), to a destination quartile bin in period one, six and 12 months ahead.

Table D1: One Step Transition Probability Matrix, 1 month ahead. Means.

	Destination bin			
Initial bin	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.862	0.093	0.025	0.020
$QB2_t$	0.084	0.796	0.092	0.029
$QB3_t$	0.018	0.091	0.793	0.096
$QB4_t$	0.007	0.014	0.070	0.911

Table D2: One Step Transition Probability Matrix, 12 months ahead. Means.

	Destination bin			
Initial bin	$QB1_{t+12}$	$QB2_{t+12}$	$QB3_{t+12}$	$QB4_{t+12}$
$QB1_t$	0.604	0.233	0.093	0.071
$QB2_t$	0.181	0.496	0.224	0.091
$QB3_t$	0.060	0.203	0.498	0.237
$QB4_t$	0.031	0.063	0.184	0.726

Table D3: One Step Transition Probability Matrix, 6 months ahead. Median, mean and standard deviation.

	Destination bin			
Initial bin	$QB1_{t+6}$	$QB2_{t+6}$	$QB3_{t+6}$	$QB4_{t+6}$
$QB1_t$	0.717	0.176	0.046	0.020
	0.685	0.195	0.070	0.051
	(0.167)	(0.109)	(0.072)	(0.102)
$QB2_t$	0.152	0.609	0.162	0.037
	0.162	0.580	0.184	0.070
	(0.091)	(0.182)	(0.112)	(0.110)
$QB3_t$	0.031	0.173	0.614	0.157
	0.047	0.177	0.580	0.193
	(0.051)	(0.081)	(0.180)	(0.145)
$QB4_t$	0.012	0.030	0.146	0.801
	0.021	0.041	0.153	0.788
	(0.025)	(0.039)	(0.068)	(0.104)

E Conditional One Step Transition Probability Matrices

The conditional TPM is based on firms that have a nominal price change.

Table E1: One Step Conditional Transition Probability Matrix, 1 month ahead. Means.

	Destination bin			
Initial bin	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.692	0.174	0.076	0.057
$QB2_t$	0.150	0.578	0.191	0.079
$QB3_t$	0.057	0.162	0.582	0.194
$QB4_t$	0.028	0.050	0.144	0.780

Table E2: One Step Conditional Transition Probability Matrix, 12 month ahead. Means.

	Destination bin			
Initial bin	$QB1_{t+12}$	$QB2_{t+12}$	$QB3_{t+12}$	$QB4_{t+12}$
$QB1_t$	0.564	0.233	0.114	0.081
$QB2_t$	0.217	0.426	0.237	0.104
$QB3_t$	0.093	0.221	0.443	0.231
$QB4_t$	0.052	0.093	0.211	0.647

Table E3: One Step Conditional Transition Probability Matrix, 6 month ahead. Median, mean and standard deviation.

	Destination bin			
Initial bin	$QB1_{t+6}$	$QB2_{t+6}$	$QB3_{t+6}$	$QB4_{t+6}$
$QB1_t$	0.648	0.200	0.071	0.030
	0.627	0.207	0.096	0.065
	(0.174)	(0.104)	(0.090)	(0.108)
$QB2_t$	0.195	0.492	0.199	0.053
	0.207	0.481	0.212	0.089
	(0.113)	(0.176)	(0.118)	(0.109)
$QB3_t$	0.061	0.209	0.503	0.187
	0.080	0.208	0.495	0.210
	(0.077)	(0.096)	(0.167)	(0.137)
$QB4_t$	0.025	0.059	0.188	0.703
	0.038	0.074	0.190	0.698
	(0.042)	(0.062)	(0.086)	(0.130)

F Duration

Table F1: Fraction of stores within each quartile bin with monthly durations (mean, median and standard deviation).

Dur.	$QB1$	$QB2$	$QB3$	$QB4$	Dur.	$QB1$	$QB2$	$QB3$	$QB4$
1 mo.	0.208	0.244	0.244	0.175	7 mo.	0.030	0.034	0.033	0.029
	0.193	0.218	0.214	0.160		0.022	0.027	0.028	0.024
	0.111	0.124	0.131	0.095		0.033	0.035	0.031	0.027
2 mo.	0.116	0.151	0.150	0.102	8 mo.	0.031	0.028	0.029	0.028
	0.105	0.142	0.139	0.093		0.024	0.022	0.022	0.023
	0.074	0.085	0.085	0.062		0.033	0.030	0.032	0.030
3 mo.	0.095	0.110	0.108	0.081	9 mo.	0.025	0.025	0.026	0.026
	0.083	0.098	0.097	0.073		0.016	0.018	0.019	0.022
	0.066	0.069	0.067	0.053		0.034	0.029	0.029	0.025
4 mo.	0.077	0.085	0.087	0.070	10 mo.	0.037	0.030	0.033	0.047
	0.067	0.077	0.082	0.063		0.026	0.023	0.026	0.040
	0.059	0.051	0.052	0.046		0.041	0.033	0.035	0.039
5 mo.	0.088	0.083	0.082	0.083	11 mo.	0.022	0.025	0.024	0.024
	0.070	0.072	0.070	0.064		0.016	0.017	0.018	0.018
	0.086	0.071	0.070	0.076		0.029	0.035	0.027	0.024
6 mo.	0.050	0.052	0.054	0.052	12 mo.	0.219	0.134	0.134	0.283
	0.042	0.044	0.045	0.044		0.207	0.112	0.119	0.277
	0.043	0.041	0.036	0.043		0.144	0.118	0.113	0.149

Note: The table consists of the mean, the median and the standard deviations for each duration within a particular quartile.

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES